
Penerapan Algoritma C4.5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta

Muhammad Hamdan Sukri¹, Yopi Handrianto²

^{1,2}Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta
10450, Indonesia
e-mail: ¹17190685@bsi.ac.id, ²yopi@bsi.ac.id

Artikel Info : Diterima : 00-00-0000 | Direvisi : 00-00-0000 | Disetujui : 00-00-0000

Abstrak - Dunia pendidikan adalah dunia lahirnya para abdi negara dan instrument pembentuk kepribadian negara. Sekolah adalah tempat terpenting dimana para "pembangun negara" Indonesia harus berjuang untuk membawa negara ini ke persaingan dunia. Penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam menentukan prediksi prestasi siswa pada SMPN 51 Jakarta. Metode yang digunakan dalam penelitian ini adalah algoritma C4.5, yang merupakan salah satu algoritma pembelajaran mesin yang populer dan efektif dalam pengambilan keputusan. Algoritma C4.5 digunakan untuk membangun model prediksi prestasi siswa berdasarkan berbagai atribut yang relevan, seperti, absensi, nilai sikap, dan ekstrakurikuler. Hasil implementasi dilakukan dengan menggunakan software Data Mining Rapid Miner, dimana hasil nantinya akan membuktikan bahwa hasil analisa yang dilakukan dengan menerapkan algoritma C4.5 sesuai dengan hasil implementasi. Hasil yang didapatkan adalah atribut "Ekstra Kulikuler" yang menjadi daun atau leaf. Kemudian pada atribut "Alpha" yang mendapatkan nilai gain tertinggi dengan hasil 1. Maka "Alpha" menjadi node akar B terakhir. Kemudian pada atribut "Alpha" SB menjadi node akar terakhir yang mendapatkan nilai gain tertinggi dengan hasil 0,468995594. Maka pada atribut "Alpha" dengan nilai >0 memiliki jawaban tidak berprestasi atau "NO" dan dengan nilai <=0 memiliki jawab berprestasi.

Kata Kunci : Prestasi Siswa, RapidMiner, Algoritma C4.5

Abstracts - The world of education is the birthplace of state servants and the instrument for forming the personality of the state. Schools are the most important places where Indonesia's "state builders" must strive to bring this country to world competition. This study aims to apply the C4.5 algorithm in determining student achievement predictions at SMPN 51 Jakarta. The method used in this study is the C4.5 algorithm, which is one of the most popular and effective machine learning algorithms for decision making. Algorithm C4.5 is used to build a predictive model of student achievement based on various relevant attributes, such as attendance, attitude scores, and extracurriculars. The results of the implementation are carried out using the Data Mining Rapid Miner software, where the results will later prove that the results of the analysis carried out by applying the C4.5 algorithm are in accordance with the results of the implementation. The results obtained are the "Extra Curricular" attributes which become leaves or leaves. Then the attribute "Alpha" gets the highest gain value with a result of 1. Then "Alpha" becomes the last B root node. Then the attribute "Alpha" SB becomes the last root node that gets the highest gain value with a result of 0.468995594. Then the attribute "Alpha" with a value > 0 has an unachieved answer or "NO" and with a value <= 0 has an achievement answer.

Keywords : Student Achivement, RapidMiner, Algorithm C4.5

PENDAHULUAN

Dunia pendidikan adalah dunia lahirnya para abdi negara dan instrument pembentuk kepribadian negara. Sekolah adalah tempat terpenting dimana para "pembangun negara" Indonesia harus berjuang untuk membawa negara ini ke persaingan dunia. Seiring dengan kesulitan dunia yang serba cepat, kesulitan dunia pendidikan semakin meningkat, perspektif ini mendorong siswa untuk melakukan yang terbaik.



Keberadaan teknologi informasi dalam dunia pendidikan dapat dengan mudah mengelola data siswa untuk menentukan prediksi prestasi Siswa. Dalam data mining dapat digunakan sebagai alat untuk pengolahan data dan dapat memberikan beberapa informasi.

SMPN 51 Jakarta merupakan salah satu lembaga pendidikan negeri daerah di Jakarta Timur. Saat ini masih banyak kondisi pemrosesan secara manual. Setiap proses yang berjalan harus menunggu proses yang sebenarnya layak dan cukup menghabiskan waktu yang relatif lama. Oleh karena itu diperlukan sistem yang dapat menghasilkan suatu sistem keputusan mengelola atau menentukan prediksi prestasi siswa. Salah satunya adalah menggunakan teknik data mining. Dalam metode Algoritma C4.5 ini adalah solusi yang dapat memudahkan sekolah dalam mengolah data dalam menentukan prestasi siswa.

Banyak metode pengolahan data berbeda-beda tergantung pada jenis data yang akan diolah, tetapi pemrosesan data yang efektif dan efisien menghasilkan informasi yang akurat. Pemrosesan data adalah elemen penting dalam teknologi informasi (Liantoni, 2022).

Penulis memilih menggunakan Algoritma C4.5 karena dapat melakukan prediksi dengan memberikan tingkat nilai akurasi yang ideal untuk memprediksi kemampuan siswa (Pambudi et al., 2018). Dimana algoritma C4.5 merupakan algoritma klasifikasi data dengan teknik pohon keputusan yang dapat mengolah data numerik (kontinyu) dan diskrit, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma lainnya.

Tinjauan Pustaka

1. Data Mining

Proses menggali nilai tambahan dari kumpulan data yang sebelumnya tidak diketahui dikenal sebagai data mining. Salah satu metode terkenal dalam data mining adalah Algoritma C4.5. Algoritma ini digunakan untuk prediksi menggunakan teknik pohon keputusan dan dapat mengolah data numerik dan diskrit. Algoritma C4.5 juga memiliki kemampuan untuk menangani nilai atribut yang hilang dan menghasilkan aturan yang mudah dipahami. Dalam konteks prediksi nilai akhir siswa, Algoritma Decision Tree dapat mencapai akurasi sekitar 85%. Namun, dengan mengoptimasi Decision Tree, tingkat akurasi untuk memprediksi siswa berpotensi bermasalah dapat meningkat menjadi 99,08%. Untuk memprediksi prestasi siswa dengan *decision tree*, diperoleh hasil akurasi yang diperoleh sangat baik yaitu 97,22%. Karena itu, penelitian ini akan using metode klasifikasi dengan Algoritma Decision Tree (Nurjana et al., 2022).

Menurut (Wahono & Riana, 2020) Data mining dibagi menjadi beberapa metode berdasarkan tugas yang dapat dilakukan:

a. Deskripsi

Deskripsi bertujuan untuk mengidentifikasi pola data yang berulang dan mengubahnya menjadi standar serta aturan yang dapat dimengerti oleh para ahli dalam bidang aplikasinya. Aturan tersebut harus mudah dipahami agar efektif meningkatkan tingkat pengetahuan sistem. Teknik *postprocessing* sering memerlukan tugas deskriptif untuk memvalidasi dan memberikan penjelasan tentang hasil dari proses Data Mining. Proses *postprocessing* digunakan untuk memastikan bahwa hanya pihak yang berkepentingan yang berhak atas hasil yang sah dan bermanfaat.

b. Prediksi

Klasifikasi dan Prediksi sebanding. Namun demikian, Data dikategorikan berdasarkan nilai atau perilaku masa depan. Tugas prediksi termasuk memprediksi jumlah pelanggan yang akan berkurang dalam waktu dekat dan memperkirakan harga saham dalam tiga bulan ke depan.

c. Estimasi

Prediksi dan estimasi serupa, namun perbedaannya terletak pada variabel target yang diestimasi. Prediksi umumnya mengacu pada peramalan nilai kategori atau diskrit, sedangkan estimasi lebih berfokus pada nilai numerik. Dalam konteks ini, nilai variabel target digunakan sebagai prediksi dalam model berdasarkan data lengkap. Kemudian, nilai variabel prediksi digunakan untuk mengestimasi nilai variabel target dalam peninjauan berikutnya.

d. Klasifikasi

Menemukan model atau fungsi yang menjelaskan dan membagi data ke dalam kelas tertentu dikenal sebagai klasifikasi. Klasifikasi adalah proses menilai karakteristik sesuatu dan memasukkannya ke dalam salah satu kelas yang sudah didefinisikan sebelumnya.

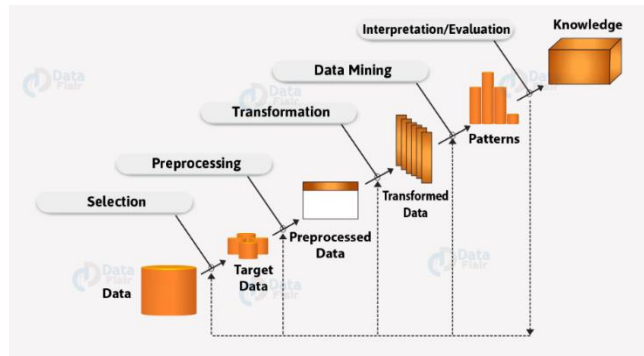
e. Clustering

Clustering adalah proses pengelompokan data ke dalam kelas objek yang serupa tanpa tergantung pada kelas data tertentu. Kluster merupakan himpunan data yang tidak hanya memiliki karakteristik yang mirip satu sama lain, tetapi juga berbeda dari data dalam kluster lain. Tujuannya adalah untuk menghasilkan

kelompok barang yang memiliki kesamaan. Kualitas analisis kluster dinilai berdasarkan tingkat kemiripan antara objek dalam satu kluster dan perbedaan antara kluster yang berbeda.

f. Asosiasi

Tugas asosiasi dalam Data Mining mencakup penemuan fitur yang muncul secara berbarengan, dengan fokus pada analisis keranjang belanja dalam konteks bisnis. Salah satu tujuannya adalah menemukan aturan yang menggambarkan hubungan antara beberapa fitur. Secara umum, Data Mining digunakan untuk tujuan deskriptif dan prediktif. Dalam deskriptif, pola-pola data yang dapat dipahami manusia dicari untuk menjelaskan karakteristik data. Sedangkan dalam prediktif, Data Mining digunakan untuk membangun model pengetahuan yang dapat digunakan untuk melakukan prediksi.



Sumber: (Muhamad et al., 2019)

Gambar 1. *Knowledge Discovery in Databases*

Tahapan data mining dibagi menjadi 6 bagian, yaitu:

a. Seleksi Data (*Data Selection*)

Database memiliki banyak data yang tidak digunakan; oleh karena itu, hanya data yang tepat untuk dianalisis yang akan diambil dari database. Sebagai contoh, dalam analisis basket pasar, tidak perlu mengambil name pelanggan; cukup menggunakan identitas pelanggan.

b. Pembersihan Data (*Data Cleaning*)

Bersihkan data melibatkan menghilangkan empat duplikat data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan data, seperti kesalahan cetak. Selain itu, proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

c. Transformasi Data (*Data Transformation*)

Transformasi data adalah proses membagi data menjadi beberapa interval karena data adalah kumpulan angka numerik yang berlanjut.

d. Proses Mining

Proses Mining adalah sebuah proses yang paling utama pada saat metode diterapkan untuk mencari pengetahuan tersembunyi dan berharga dari data.

e. Evaluasi Pola (*Pattern Evaluation*)

Pada titik ini, hasil dari metode data mining, termasuk pola-pola khusus dan model prediksi, dievaluasi untuk mengetahui apakah hipotesa saat ini tercapai.

f. Presentasi Pengetahuan (*Knowledge Presentation*)

Salah satu tahapan yang diperlukan dalam proses Data Mining adalah menyampaikan hasilnya dalam bentuk pengetahuan yang dapat dipahami oleh semua orang. Visualisasi juga dapat membantu menyampaikan hasil Data Mining dalam presentasi ini.

2. Klasifikasi

Klasifikasi adalah tugas yang sangat sering terjadi dalam kehidupan sehari-hari. Pada dasarnya, item dibagi sehingga masing-masing diberi kelas, yaitu nomor kategori yang saling eksklusif. Itu hanya berarti bahwa setiap objek hanya boleh ditugaskan ke satu kelas, dan tidak pernah ada lebih dari satu atau kelas sama sekali (Yendrizal, 2022).

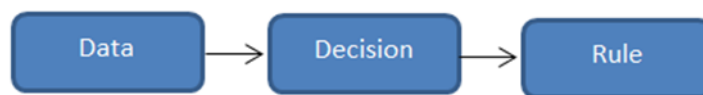
Tujuan dari klasifikasi adalah menemukan model yang dapat mengkategorikan atribut kedalam kategori atau kelas yang sesuai. Tujuan dari metode ini adalah untuk menentukan kategori dari objek yang modelnya belum diketahui. ID3, CART, dan C4.5 adalah beberapa teknik klasifikasi.

3. Pohon Keputusan

Pohon keputusan adalah teknik yang efektif untuk klasifikasi dan prediksi. Metodenya menghasilkan pohon keputusan yang mewakili aturan dari data dengan bahasa alami yang mudah dipahami. Pohon keputusan dapat berinteraksi dengan data dalam bahasa yang digunakan dalam basis data, seperti Bahasa Pertanyaan Struktural. Mereka berguna dalam analisis data untuk mengungkap hubungan antara variabel target dan variabel input. Pohon keputusan menggabungkan eksplorasi data dan pemodelan, sering digunakan sebagai langkah awal dalam proses pemodelan atau sebagai model akhir.

Teknik klasifikasi yang paling kreatif adalah pohon keputusan, yang terdiri dari node pilihan yang terhubung oleh cabang dari node akar menuju node daun (ujung). Pilihan node akan dicoba dalam hal kualitas, dan setiap hasil akan membuat cabang. Selanjutnya, cabang tersebut akan dihubungkan ke node lain atau node akhir untuk membuat keputusan (Kusrini & Luthfi, 2019).

Biasanya, pohon keputusan digunakan untuk mendapatkan informasi yang bertujuan untuk meningkatkan proses pengambilan keputusan. Pohon keputusan diawali dengan root node, atau titik kontrol, yang digunakan oleh pengguna untuk membuat keputusan. User memecahkan leaf node berdasarkan root node ini sesuai dengan Algoritma *Decision Tree* (Ramadhani & Hendriyani, 2021).



Gambar 2. Alur *Decision Tree*

4. Algoritma C4.5

Pohon Keputusan adalah hasil dari Algoritma C4.5, yang merupakan salah satu Algoritma klasifikasi yang sangat efektif dan sering digunakan dalam berbagai konteks. Algoritma C4.5 adalah pengembangan dari Algoritma Iterative Dichotomiser 3 (ID3) yang diperkenalkan oleh J. Ross Quinlan pada tahun 1996. Algoritma C4.5 menghasilkan sistem pembentukan pohon keputusan yang berhasil berkat perbaikan dari algoritma ID3. Beberapa perbaikan termasuk pemangkasan pohon, menangani hilangnya data, dan menangani data kontinu (Najib et al., 2019).

Keuntungan utama Algoritma C4.5 dapat digunakan untuk menghasilkan Pohon Keputusan (*Decision Tree*) yang efisien yang menangani atribut numerik tipe diskrit and tipe diskrit, yang mudah diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima. Dalam (Handrianto & Farhan, 2019) langkah-langkah pembuatan pohon keputusan Algoritma C4.5 ialah:

- a. Menyiapkan Data Pelatihan. Data Pelatihan biasanya diambil dari data historis yang telah terjadi sebelumnya, atau disebut sebagai data masa lalu, dan dikelompokkan ke dalam kelas-kelas yang berbeda.
- b. Menghitung akar pohon Akar diambil dari atribut yang dipilih dengan menghitung nilai gain. Nilai pasir tertinggi dari setiap atribut adalah akar pertama. Sebelum menghitung nilai gain atribut, terlebih dahulu menghitung nilai entropy.
- c. Ulangi langkah 2 dan 3 sampai semua rekaman ditutup.
- d. Proses partisi pohon keputusan berhenti ketika:
 - 1) Semua record dari node N memiliki class yang sama.
 - 2) Catatan yang dipartisi tidak lagi memiliki atribut.
 - 3) Cabang kosong tidak memiliki catatan.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.

Untuk menghitung gain digunakan rumus seperti yang tertera berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropy. Entropy digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

S : Himpunan Kasus

n : Jumlah partisi S

pi: Proporsi dari Si terhadap S

5. Prestasi Belajar

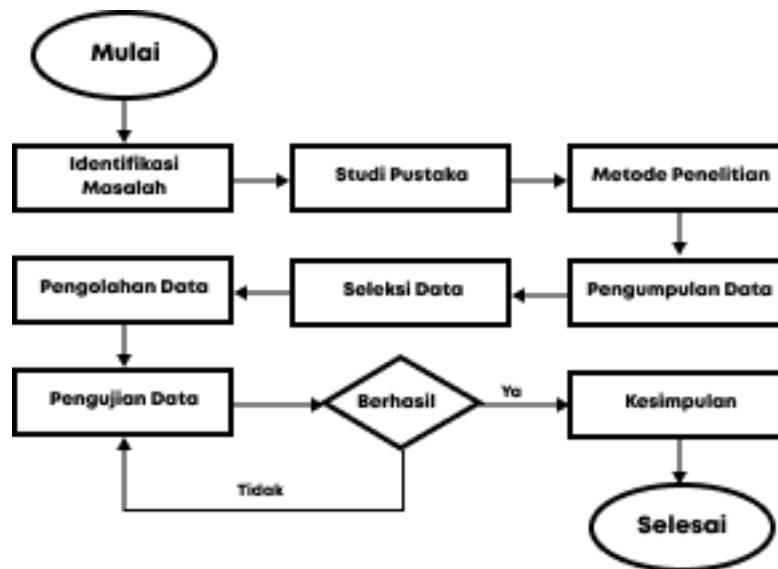
Keberhasilan dalam membentuk siswa yang berkualitas tinggi adalah fokus utama proses belajar mengajar, dan keberhasilan ini sangat berkaitan dengan kualitas pendidikan. Dengan belajar, siswa diharapkan memperoleh banyak pengetahuan dan wawasan, yang merupakan bagian penting dari keberhasilan proses pembelajaran (Rosyid et al., 2020).

Prestasi adalah ukuran hasil pendidikan yang diukur berdasarkan penggunaan istilah dan tata bahasa yang benar dalam Kamus Besar Bahasa Indonesia. Prestasi memungkinkan seseorang menilai kinerjanya dengan standar internal dan eksternal serta bersaing dengan orang lain. Siswa diartikan sebagai anak yang sedang belajar dalam Kamus Besar Bahasa Indonesia, dan setiap siswa memiliki potensi dan pengalaman pertumbuhan yang berbeda. Guru tidak menentukan karakter dan gaya hidup anak, tetapi interaksi sosial anak dan dirinya sendiri yang menentukannya. Pelajar adalah individu yang berada di pusat proses belajar-mengajar, memiliki tujuan, dan berusaha mencapainya dengan cara terbaik.

Prestasi siswa adalah hasil belajar dalam bentuk keterampilan, kemampuan, dan pengetahuan yang digunakan. Nilai akhir semester bisa mencerminkan keberhasilan belajar, tetapi bisa juga dipengaruhi oleh masalah yang menghambat pembelajaran. Prestasi siswa mencakup berbagai aspek seperti pengetahuan, pemahaman, keterampilan, sikap, dan perilaku, serta proses belajar-mengajar. Kegiatan ini juga bisa memengaruhi kecerdasan siswa dan efisiensinya dalam belajar. Perubahan dalam perilaku siswa dapat memengaruhi prestasi mereka, dan dorongan positif dapat membantu siswa tampil dengan baik dalam pendidikan.

METODE PENELITIAN

1. Tahapan Penelitian



Sumber: Peneliti 2023

Gambar 3. Tahapan Penelitian

Secara umum, berikut adalah penjelasan tentang metodologi penelitian :

- Identifikasi Masalah
Memanfaatkan literatur dan data yang telah diperoleh untuk menentukan masalah manajemen kualitas dan kegagalan konstruksi yang akan dibahas.
- Melakukan Studi Pustaka
Untuk memulai, pertama-tama dilakukan penelitian literatur terkait dengan studi sebelumnya tentang penggunaan Algoritma C4.5 dan kemudian menguatkan penelitian ini berdasarkan teori yang digunakan. Dalam contoh ini, peneliti ingin menganalisis Prediksi Prestasi Siswa SMPN 51 Jakarta dengan

menggunakan Algoritma C4.5. Sekaligus juga ingin mempelajari pola dan mekanisme klasifikasi dalam data mining.

c. Metode Penelitian

Setelah studi pustaka selesai, langkah selanjutnya adalah memilih Metode Klasifikasi. Setelah melakukan pengamatan dan observasi, peneliti membuat keputusan tentang metode untuk klasifikasi. Karena Algoritma C4.5 memiliki kemampuan untuk melakukan Klasifikasi Data dengan membuat Pohon Keputusan dan Mengolah Data dalam bentuk numerik dan kategorikal.

d. Pengumpulan Data

Proses ini dilakukan dengan mendatangi pihak instansi yang relevan secara langsung untuk mendapatkan data yang diperlukan.

e. Seleksi Data

Pada tahap ini, data *database* dipilih. Karena data yang diperoleh tidak semuanya digunakan, seleksi data dilakukan untuk memilih atribut atau *variabel* yang dibutuhkan dalam penelitian. Pada penelitian ini, Izin, Alpha, Sakit, Sikap, Pramuka wajib, dan Extra Kurikuler adalah atribut yang dipilih. Pada langkah ini, data *null*, tidak valid, dan ganda akan dihilangkan. Karena data yang tidak lengkap atau tidak valid akan memengaruhi hasil yang dihasilkan. Selanjutnya, *Microsoft Office Excel 2007* akan digunakan untuk mengolah data.

f. Pengolahan Data

Setelah semua data yang diperlukan telah dipilih, tahap penelitian berikutnya adalah pengolahan data. Pada tahap ini, nilai atribut data akan diubah kedalam bentuk data yang sesuai untuk memungkinkan pemrosesan data Algoritma C4.5. untuk mendapatkan *dataset* utuh yang akan digunakan untuk melanjutkan proses.

g. Pengujian Data

Pada tahap pengujian hasil, pengujian data akan dilakukan baik secara manual dengan algoritma C4.5 maupun dengan *Software RapidMiner*. Berikut adalah langkah-langkah untuk menggunakan *Program RapidMiner*:

- 1) Untuk memulai tahapan pengujian dengan *RapidMiner*, pertama-tama buka aplikasinya. Kemudian, klik *New Process*, dan Anda akan melihat tampilan proses utama, yang akan digunakan untuk melakukan proses baru.
- 2) Setelah itu, klik *Import* di *Panel Operators*, pilih *Data*, dan pilih *Read Excel*. Tampilan *Read Excel* akan muncul di *Main Process*.
- 3) Setelah itu, klik *Import Configuration Wizard* untuk memasukkan *dataset Excel* yang telah dibuat, *Configuration Wizard* akan menampilkan empat tahapan : langkah pertama adalah memilih informasi yang akan digunakan, pada Tahap kedua adalah memilih lembar kerja untuk memasukkan *dataset*, pada tahap ketiga mengisi *Annotation* jika diperlukan, dan pada tahap keempat mengubah atribut sesuai kebutuhan. Kemudian klik *Selesai*.
- 4) Kemudian, klik *Modeling* di *Panel Operators* dan pilih *Klasifikasi* dan *Regression*, yang dapat diubah jika diperlukan. Metode ini dipilih karena penelitian ini akan menggunakan kategorisasi. Klik *Operator Baru*, pilih *Induksi Pohon*, dan kemudian klik *Pohon Keputusan*. sebagai metode yang akan menentukan hasil akhirnya.
- 5) Selanjutnya, *Port Read Excel* dihubungkan ke *Port Decision Tree*. Setelah semua terhubung, klik *Run* untuk melihat hasil dari semua proses. Tahapan penelitian kembali ke pengolahan data jika pengujian tidak menghasilkan output yang memenuhi persyaratan.

Cara pengujian yang dilakukan menggunakan metode Algoritma C4.5 adalah sebagai berikut:

- 1) Pertama, dataset untuk pengujian disiapkan.
- 2) Selanjutnya, gunakan rumus berikut untuk menghitung nilai *Entropy* dan *Information Gain* dari jumlah data tersebut:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Digunakan untuk menghitung nilai entropy dari semua atribut yang ada dalam dataset.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Digunakan untuk menghitung nilai *Gain* untuk masing-masing atribut dataset.

$$Akurasi = \frac{Jumlah\ Prediksi}{Jumlah\ Total} * 100\%$$

Digunakan untuk menghitung nilai Akurasi.

- 3) Kemudian membuat pohon keputusan dengan menghitung nilai gain yang paling tinggi dari setiap atribut yang ada. Nilai *gain ratio* tertinggi dari atribut tersebut akan digunakan sebagai akar dalam pohon keputusan.

- 4) Selanjutnya, buat cabang dari nilai atribut yang tersisa sesuai dengan jumlah nilai rasio peningkatan variabel. secara efektif menggunakan rumus *Entropy (S)* dan *Gain (S,A)*.
- 5) Terakhir, proses diulangi pada setiap cabang. sampai semua kasus memiliki kelas yang sama dan proses tidak dapat dilakukan lagi.
- h. Membuat Kesimpulan
Keputusan dibuat berdasarkan analisis data, yang kemudian diperiksa untuk memastikan bahwa hasilnya sesuai dengan tujuan dan maksud penelitian

2. Metode Pengumpulan Data

Penelitian ini menggunakan metode Studi Pustaka. Untuk mendapatkan data yang dibutuhkan, teknik ini digunakan untuk mencari informasi tentang pengolahan data dari Buku, E-Book , Jurnal Ilmiah, dan Literatur lainnya. Data yang digunakan dalam penelitian ini adalah sebagai berikut:

- a. Data Primer
Data primer diambil secara langsung dari data siswa di SMPN 51 Jakarta yang berhubungan dengan penelitian yang dilakukan dengan cara Observasi dan Wawancara.
 - 1) Observasi
Observasi dilakukan untuk mendapatkan pemahaman tentang suatu peristiwa atau peristiwa secara langsung untuk menjawab pertanyaan penelitian (Hermawan, 2019).
 - 2) Wawancara
Peneliti melakukan wawancara dengan subjek penelitian atau informan untuk mendapatkan informasi. Pada dasarnya, wawancara adalah cara untuk mendapatkan informasi lebih lanjut tentang subjek penelitian. Atau, itu adalah proses membuktikan informasi atau keterangan yang telah diperoleh dari metode sebelumnya. Wawancara digunakan sebagai metode pengumpulan data apabila peneliti ingin melakukan studi pendahuluan untuk menemukan masalah yang harus diteliti, serta apabila peneliti ingin mengetahui lebih banyak tentang responden (Hermawan, 2019).
- b. Data Sekunder
Data Sekunder digunakan untuk melengkapi data primer yang dikumpulkan from buku-buku, jurnal ilmiah, publikasi, laporan penelitian, dan internet untuk mendukung landasan teori.
Dengan metode ini, tugas yang dilakukan adalah mencari dan mengumpulkan data yang berkaitan dengan prediksi prestasi siswa yang diperoleh dari SMPN 51 Jakarta. Selanjutnya, data yang telah diperoleh akan diolah menggunakan metode Klasifikasi Algoritma C4.5. Metode ini akan memungkinkan untuk menentukan prediksi prestasi siswa dengan data yang dikumpulkan. Adapun pengumpulan data yang di ambil, yaitu :
 - 1) Pengambilan Populasi
Peneliti ingin meneliti semua komponen data yang tersedia dalam data yang diselidiki, jadi penelitiannya melibatkan populasi. Data penelitian berasal dari Siswa SMPN 51 Jakarta. Data populasi terdiri dari sampel. Sebuah populasi dapat diwakili dengan pemilihan tertentu yang merupakan bagian dari sample terbaik. Akhirnya, data yang diolah dan dijadikan sampel adalah siswa di SMPN 51 Jakarta. Data seluruh siswa kelas IX digunakan untuk populasi penelitian ini sebagai data kelulusan Siswa SMPN 51 Jakarta.
 - 2) Pengambilan Sampel
Sample penelitian dipilih untuk mengumpulkan informasi tentang subjek penelitian. Representasi dari populasi yang diproyeksikan dapat mewakili seluruh populasi dipilih. Penulis akan menggunakan sample berikut:

Tabel 1. Sampel Penelitian

KELAS	JUMLAH SAMPEL
IX 1	38
IX 2	39
IX 3	40
IX 4	40
IX 5	39
IX 6	40
IX 7	40
IX 8	40
JUMLAH	316

Sumber: Peneliti 2023

3. Metode Analisis Data

Menurut (Sugiyono, 2015) Penelitian kuantitatif, berdasarkan Filsafat Positiveme, digunakan untuk mempelajari populasi atau sampel tertentu. Pengambilan sampel biasanya dilakukan secara acak dan data

dikumpulkan menggunakan instrumen penelitian. Tujuan dari analisis kuantitatif dan statistik data adalah untuk menguji hipotesis yang telah ditetapkan. Filsafat Positivisme melihat realitas yang dapat diklasifikasikan, relatif tetap, konkrit, teramati, terukur, dan hubungan gejala sampel tertentu yang representatif, Pengambilan data akan diatur oleh kriteria yang akan diteliti, yaitu:

- Sampel Data Siswa SMPN 51 Jakarta.
- Menentukan siswa yang berprestasi atau tidak dengan membuat pohon keputusan melalui rumus *Entropy*, *Gain*, dan mencari rata-rata *akurasi*.
- Mengolah data dengan *Rapid Miner*.

HASIL DAN PEMBAHASAN

1. Hasil Klasifikasi

Dalam penelitian ini Metode yang digunakan adalah Algoritma C4.5. Langkah pertama adalah menghitung nilai gain dan entropy. Perhitungan nilai gain menggunakan data pelatihan, yaitu lima atribut dari 316 data yang diambil. Tujuan penggunaan data sampel adalah untuk membuat proses menghitung nilai gain untuk setiap atribut lebih mudah. Ini adalah batasan untuk atribut yang diambil:

- Sakit (>3 dan ≤ 3)
- Izin (>0 dan ≤ 0)
- Alpha (>0 dan ≤ 0)
- Sikap (A, B dan C)
- Pramuka Wajib (SB, B dan C)
- Extra Kurikuler (SB, B dan Tidak Ikut)

1.1. Perhitungan Manual Algoritma C4.5

Langkah pertama untuk menggunakan data pelatihan sebanyak 316 data siswa dengan menggunakan Algoritma C4.5 untuk menentukan pohon keputusan adalah membuat 316 data untuk penelitian dalam data pelatihan. Data pelatihan telah dikelompokkan ke dalam kelas tertentu yang diperoleh dari data primer

Tabel 2. *Data Training*

No	Nama Siswa	Sakit	Izin	Alpha	Sikap	Pramuka Wajib	Extra Kurikuler	Prestasi
1	Aisya Razma Salimah	≤ 3	≤ 0	≤ 0	B	B	Tidak Ikut	No
2	Andreas Octaviano Simanjuntak	≤ 3	≤ 0	≤ 0	B	B	Tidak Ikut	No
3	Callista Sunday Zalianov	≤ 3	≤ 0	≤ 0	B	B	Tidak Ikut	No
4	Faiz Ardiansyah	≤ 3	≤ 0	≤ 0	B	B	Tidak Ikut	No
5	Fathan Akbar Tristanto	> 3	≤ 0	≤ 0	B	B	Sb	Yes
6	Fauzan Ramadhan Rafilis	> 3	≤ 0	> 0	B	B	Tidak Ikut	No
7	Fitri Nurandini	≤ 3	≤ 0	≤ 0	B	B	Sb	Yes
8	Ghana Luthfi Syawali	≤ 3	≤ 0	≤ 0	A	B	B	Yes
9	Gilang Permadi Putra	> 3	≤ 0	≤ 0	B	B	Tidak Ikut	No
10	Gisela Christa Parura	≤ 3	≤ 0	≤ 0	A	B	B	Yes

Dan seterusnya.

Sumber: Peneliti 2023

Tabel 3. Hasil Perhitungan Node 1

Node	Atribut	Value	Jumlah Kasus	Yes	No	Entropy	Gain
1	Total		316	108	208	0,926504446	

							0,030536131
	Sakit	>3	65	10	55	0,619382195	
		<=3	251	99	152	0,967594202	
							0,029806122
	Izin	>3	56	9	50	0,569853677	
		<=3	260	99	158	0,967095632	
	Alpha	>3	33	0	33	0	0,065320828
		<=3	283	109	174	0,961604322	
	Sikap	A	65	35	30	0,995727452	0,0283752
		B	250	74	176	0,876346229	
		C	1	0	1	0	
	Pramuka Wajib	Sb	25	10	15	0,970950594	-0,00024299
		B	290	99	191	0,926132498	
		C	1	0	1	0	
	Ekstra Kulikuler	Sb	73	66	7	0,455831458	0,70497929
		B	52	42	10	0,706274089	
		Tidak Ikut	191	0	191	0	

Sumber: Peneliti 2023

$$\text{Entropy Total} = \left(\left(-\frac{108}{316} \right) \times \log_2 \left(\frac{108}{316} \right) \right) + \left(\left(-\frac{208}{316} \right) \times \log_2 \left(\frac{208}{316} \right) \right) = 0,926504446$$

$$\text{Entropy Sakit (>3)} = \left(\left(-\frac{10}{65} \right) \times \log_2 \left(\frac{10}{65} \right) \right) + \left(\left(-\frac{55}{65} \right) \times \log_2 \left(\frac{55}{65} \right) \right) = 0,619382195$$

$$\text{Entropy Sakit (<=3)} = \left(\left(-\frac{99}{251} \right) \times \log_2 \left(\frac{99}{251} \right) \right) + \left(\left(-\frac{152}{251} \right) \times \log_2 \left(\frac{152}{251} \right) \right) = 0,967594202$$

$$\text{Entropy Izin (>3)} = \left(\left(-\frac{9}{56} \right) \times \log_2 \left(\frac{9}{56} \right) \right) + \left(\left(-\frac{50}{56} \right) \times \log_2 \left(\frac{50}{56} \right) \right) = 0,569853677$$

$$\text{Entropy Izin (<=3)} = \left(\left(-\frac{99}{260} \right) \times \log_2 \left(\frac{99}{260} \right) \right) + \left(\left(-\frac{158}{260} \right) \times \log_2 \left(\frac{158}{260} \right) \right) = 0,967095632$$

$$\text{Entropy Alpha (>3)} = \left(\left(-\frac{0}{33} \right) \times \log_2 \left(\frac{0}{33} \right) \right) + \left(\left(-\frac{33}{33} \right) \times \log_2 \left(\frac{33}{33} \right) \right) = 0$$

$$\text{Entropy Alpha (<=3)} = \left(\left(-\frac{109}{283} \right) \times \log_2 \left(\frac{109}{283} \right) \right) + \left(\left(-\frac{174}{283} \right) \times \log_2 \left(\frac{174}{283} \right) \right) = 0,961604322$$

$$\text{Entropy Sikap (A)} = \left(\left(-\frac{35}{65} \right) \times \log_2 \left(\frac{35}{65} \right) \right) + \left(\left(-\frac{30}{65} \right) \times \log_2 \left(\frac{30}{65} \right) \right) = 0,995727452$$

$$\text{Entropy Sikap (B)} = \left(\left(-\frac{74}{250} \right) \times \log_2 \left(\frac{74}{250} \right) \right) + \left(\left(-\frac{176}{250} \right) \times \log_2 \left(\frac{176}{250} \right) \right) = 0,876346229$$

$$\text{Entropy Sikap (C)} = \left(\left(-\frac{0}{1} \right) \times \log_2 \left(\frac{0}{1} \right) \right) + \left(\left(-\frac{1}{1} \right) \times \log_2 \left(\frac{1}{1} \right) \right) = 0$$

$$\text{Entropy Pramuka Wajib (SB)} = \left(\left(-\frac{10}{25} \right) \times \log_2 \left(\frac{10}{25} \right) \right) + \left(\left(-\frac{15}{25} \right) \times \log_2 \left(\frac{15}{25} \right) \right) = 0,970950594$$

$$\text{Entropy Pramuka Wajib (B)} = \left(\left(-\frac{99}{290} \right) \times \log_2 \left(\frac{99}{290} \right) \right) + \left(\left(-\frac{191}{290} \right) \times \log_2 \left(\frac{191}{290} \right) \right) = 0,926132498$$

$$\text{Entropy Pramuka Wajib (C)} = \left(\left(-\frac{0}{1} \right) \times \log_2 \left(\frac{0}{1} \right) \right) + \left(\left(-\frac{1}{1} \right) \times \log_2 \left(\frac{1}{1} \right) \right) = 0$$

$$\text{Entropy Ekstra Kulikuler (SB)} = \left(\left(-\frac{66}{73} \right) \times \log_2 \left(\frac{66}{73} \right) \right) + \left(\left(-\frac{7}{73} \right) \times \log_2 \left(\frac{7}{73} \right) \right) = 0,455831458$$

$$\text{Entropy Ekstra Kulikuler (B)} = \left(\left(-\frac{42}{52} \right) \times \log_2 \left(\frac{42}{52} \right) \right) + \left(\left(-\frac{10}{52} \right) \times \log_2 \left(\frac{10}{52} \right) \right) = 0,706274089$$

$$\text{Entropy Ekstra Kulikuler (C)} = \left(\left(-\frac{0}{191} \right) \times \log_2 \left(\frac{0}{191} \right) \right) + \left(\left(-\frac{191}{191} \right) \times \log_2 \left(\frac{191}{191} \right) \right) = 0$$

setelah menghitung nilai gain dari setiap atribut. Maka hasilnya adalah atribut “ Ekstra Kulikuler ” yang mendapatkan nilai gain tertinggi dengan hasil 0,70793948. Maka “ Ekstra Kulikuler ” menjadi node akar. Kemudian pada atribut “ Ekstra Kulikuler ” dengan nilai Tidak Ikut memiliki jawaban tidak berprestasi atau “NO”. Demikian nilai Ekstra Kulikuler menjadi daun atau leaf.

Tabel 4. Hasil Perhitungan Node 2

Node	Atribut	Value	Jumlah Kasus	Yes	No	Entropy	Gain
2	Ekstra Kulikuler = (B)		52	42	10	0,706274089	0,431140797
	Sakit	>3	11	2	9	0,684038436	
		<=3	41	40	1	0,165427034	
							0,274054314
	Izin	>3	12	4	8	0,918295834	
		<=3	40	38	2	0,286396957	
							0,148736279
	Alpha	>3	3	0	3	0	
		<=3	49	42	7	0,591672779	
							0,053970463
	Sikap	A	18	17	1	0,309543429	
		B	34	25	9	0,833764907	
		C	0	0	0	0	
							0,012112459
	Pramuka Wajib	Sb	2	2	0	0	
		B	50	40	10	0,721928095	
		C	0	0	0	0	

Sumber: Peneliti 2023

$$\text{Entropy Sakit (>3)} = \left(\left(-\frac{2}{11} \right) \times \log_2 \left(\frac{2}{11} \right) \right) + \left(\left(-\frac{9}{11} \right) \times \log_2 \left(\frac{9}{11} \right) \right) = 0,684038436$$

$$\text{Entropy Sakit (<=3)} = \left(\left(-\frac{40}{41} \right) \times \log_2 \left(\frac{40}{41} \right) \right) + \left(\left(-\frac{1}{41} \right) \times \log_2 \left(\frac{1}{41} \right) \right) = 0,165427034$$

$$\text{Entropy Izin (>3)} = \left(\left(-\frac{4}{12} \right) \times \log_2 \left(\frac{4}{12} \right) \right) + \left(\left(-\frac{8}{12} \right) \times \log_2 \left(\frac{8}{12} \right) \right) = 0,918295834$$

$$\text{Entropy Izin (<=3)} = \left(\left(-\frac{38}{40} \right) \times \log_2 \left(\frac{38}{40} \right) \right) + \left(\left(-\frac{2}{40} \right) \times \log_2 \left(\frac{2}{40} \right) \right) = 0,286396957$$

$$\text{Entropy Alpha (>3)} = \left(\left(-\frac{0}{3} \right) \times \log_2 \left(\frac{0}{3} \right) \right) + \left(\left(-\frac{3}{3} \right) \times \log_2 \left(\frac{3}{3} \right) \right) = 0$$

$$\text{Entropy Alpha (<=3)} = \left(\left(-\frac{42}{49} \right) \times \log_2 \left(\frac{42}{49} \right) \right) + \left(\left(-\frac{7}{49} \right) \times \log_2 \left(\frac{7}{49} \right) \right) = 0,591672779$$

$$\text{Entropy Sikap (A)} = \left(\left(-\frac{17}{18} \right) \times \log_2 \left(\frac{17}{18} \right) \right) + \left(\left(-\frac{1}{18} \right) \times \log_2 \left(\frac{1}{18} \right) \right) = 0,309543429$$

$$\text{Entropy Sikap (B)} = \left(\left(-\frac{25}{34} \right) \times \log_2 \left(\frac{25}{34} \right) \right) + \left(\left(-\frac{9}{34} \right) \times \log_2 \left(\frac{9}{34} \right) \right) = 0,833764907$$

$$\text{Entropy Sikap (C)} = \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) + \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) = 0$$

$$\text{Entropy Pramuka Wajib (SB)} = \left(\left(-\frac{2}{2} \right) \times \log_2 \left(\frac{2}{2} \right) \right) + \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) = 0$$

$$\text{Entropy Pramuka Wajib (B)} = \left(\left(-\frac{40}{50} \right) \times \log_2 \left(\frac{40}{50} \right) \right) + \left(\left(-\frac{10}{50} \right) \times \log_2 \left(\frac{10}{50} \right) \right) = 0,721928095$$

$$\text{Entropy Pramuka Wajib (C)} = \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) + \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) = 0$$

Tabel 5. Hasil Perhitungan Node 3(<=3)

Node	Atribut	Value	Jumlah Kasus	Yes	No	Entropy	Gain
3	Sakit = (<=3)		41	40	1	0,165427034	
							0,077387022
	Izin	>3	5	4	1	0,721928095	
		<=3	36	36	0	0	
							0
	Alpha	>3	0	0	0	0	
		<=3	41	40	1	0,165427034	
							0,019154472
	Sikap	A	17	17	0	0	
		B	24	23	1	0,249882293	
		C	0	0	0	0	
							0,001782131
	Pramuka Wajib	Sb	2	2	0	0	
		B	39	38	1	0,172036949	
		C				0	

Sumber: Peneliti 2023

$$\text{Entropy Izin (>3)} = \left(\left(-\frac{4}{5} \right) \times \log_2 \left(\frac{4}{5} \right) \right) + \left(\left(-\frac{1}{5} \right) \times \log_2 \left(\frac{1}{5} \right) \right) = 0,721928095$$

$$\text{Entropy Izin (<=3)} = \left(\left(-\frac{0}{36} \right) \times \log_2 \left(\frac{0}{36} \right) \right) + \left(\left(-\frac{36}{36} \right) \times \log_2 \left(\frac{36}{36} \right) \right) = 0$$

$$\text{Entropy Alpha (>3)} = \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) + \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) = 0$$

$$\text{Entropy Alpha (<=3)} = \left(\left(-\frac{40}{41} \right) \times \log_2 \left(\frac{40}{41} \right) \right) + \left(\left(-\frac{1}{41} \right) \times \log_2 \left(\frac{1}{41} \right) \right) = 0,165427034$$

$$\text{Entropy Sikap (A)} = \left(\left(-\frac{17}{17} \right) \times \log_2 \left(\frac{17}{17} \right) \right) + \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) = 0$$

$$\text{Entropy Sikap (B)} = \left(\left(-\frac{23}{24} \right) \times \log_2 \left(\frac{23}{24} \right) \right) + \left(\left(-\frac{1}{24} \right) \times \log_2 \left(\frac{1}{24} \right) \right) = 0,249882293$$

$$\text{Entropy Sikap (C)} = \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) + \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) = 0$$

$$\text{Entropy Pramuka Wajib (SB)} = \left(\left(-\frac{2}{2} \right) \times \log_2 \left(\frac{2}{2} \right) \right) + \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) = 0$$

$$\text{Entropy Pramuka Wajib (B)} = \left(\left(-\frac{38}{39} \right) \times \log_2 \left(\frac{38}{39} \right) \right) + \left(\left(-\frac{1}{39} \right) \times \log_2 \left(\frac{1}{39} \right) \right) = 0,172036949$$

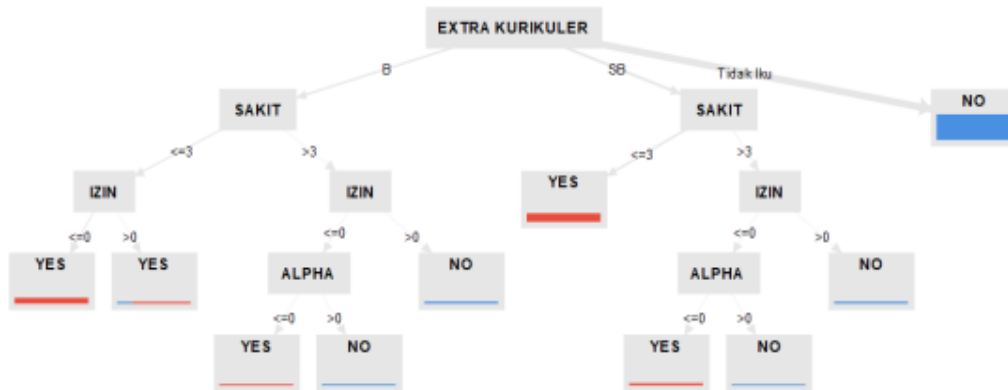
$$\text{Entropy Pramuka Wajib (C)} = \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) + \left(\left(-\frac{0}{0} \right) \times \log_2 \left(\frac{0}{0} \right) \right) = 0$$

Menghitung presentase Prestasi YES dan NO dengan menggunakan rumus secara manual :

$$\text{YES} : \frac{108}{316} \times 100\% = 34,18\%$$

$$\text{NO} : \frac{208}{316} \times 100\% = 65,82\%$$

1.2. Pohon Keputusan (*Decision Tree*)



Sumber: Peneliti 2023

Gambar 4. Tampilan *Decision Tree*

Adapun penjelasan pada pohon keputusan gambar diatas. Yang mendapatkan rule untuk menentukan prediksi prestasi siswa dapat dilihat pada text view :

Tree

```
EXTRA KURIKULER = B
|   SAKIT = <=3
|   |   IZIN = <=0: YES {NO=0, YES=36}
|   |   IZIN = >0: YES {NO=1, YES=4}
|   SAKIT = >3
|   |   IZIN = <=0
|   |   |   ALPHA = <=0: YES {NO=0, YES=2}
|   |   |   ALPHA = >0: NO {NO=2, YES=0}
|   |   IZIN = >0: NO {NO=7, YES=0}
EXTRA KURIKULER = SB
|   SAKIT = <=3: YES {NO=0, YES=57}
|   SAKIT = >3
|   |   IZIN = <=0
|   |   |   ALPHA = <=0: YES {NO=0, YES=9}
|   |   |   ALPHA = >0: NO {NO=1, YES=0}
|   |   IZIN = >0: NO {NO=6, YES=0}
EXTRA KURIKULER = Tidak Iku: NO {NO=191, YES=0}
```

Sumber: Peneliti 2023

Gambar 5. Tampilan *Rule Decision Tree*

Penjelasan Rule pada *Decision Tree*:

- Jika EXTRA KURIKULER = B , SAKIT = <=3, IZIN = <=0 maka hasilnya YES {NO=0, YES=36}
- Jika EXTRA KURIKULER = B , SAKIT = <=3, IZIN = >0 maka hasilnya YES {NO=1, YES=4}.
- Jika EXTRA KURIKULER = B, SAKIT = >3, IZIN = <=0, ALPHA = <=0 Maka hasilnya YES {NO=0, YES=2}
- Jika EXTRA KURIKULER = B, SAKIT = >3, IZIN = <=0, ALPHA = >0 maka hasilnya NO {NO=2, YES=0}
- Jika EXTRA KURIKULER = B, SAKIT = >3, IZIN = >0 maka hasilnya NO {NO=7, YES=0}
- Jika EXTRA KURIKULER = SB, SAKIT = <=3, maka hasilnya YES {NO=0, YES=57}

- g. Jika EXTRA KURIKULER SB, SAKIT = >3, IZIN = <=0, ALPHA = <=0 maka hasilnya YES {NO=0, YES=9}
- h. Jika EXTRA KURIKULER SB, SAKIT = >3, IZIN = <=0, ALPHA = >0 Maka hasilnya NO {NO=1, YES=0}
- i. Jika EXTRA KURIKULER SB, IZIN = >0 Maka hasilnya NO {NO=6, YES=0}
- j. Jika EXTRA KURIKULER = Tidak Ikut, Maka hasilnya NO {NO=191, YES=0}

2. Pembahasan

Berdasarkan penelitian diatas, beberapa point yang dapat ditarik pembahasannya ialah:

- a. Hasil dari Node 1, dapat dilihat bahwa hasilnya adalah atribut " Ekstra Kulikuler " yang mendapatkan nilai gain tertinggi dengan hasil 0,70793948. Maka " Ekstra Kulikuler " menjadi node akar. Kemudian pada atribut " Ekstra Kulikuler " dengan nilai Tidak Ikut memiliki jawaban tidak berprestasi atau "NO". Demikian nilai Ekstra Kulikuler menjadi daun atau leaf.
- b. Hasil dari Node 2, dapat dilihat bahwa hasilnya adalah atribut " Sakit " yang mendapatkan nilai gain tertinggi dengan hasil 0,431140797. Kemudian pada atribut "Sakit" mencari nilai "Sakit <=3". Menentukan simpul berikutnya yaitu Node 3, kemudian menentukan kembali nilai entropy dan gain dari nilai Sakit "<=3" untuk mendapatkan perhitungan nilai gain tertinggi.
- c. Hasil dari Node 3, dapat dilihat bahwa setelah menghitung nilai gain dari setiap atribut "Izin <=3" menyatakan bahwa hasil tersebut mendapatkan hasil nilai gain tertinggi dengan hasil 0,077387022 sekaligus menyatakan siswa yang berprestasi.
- d. Hasil penerapan Algoritma C4.5 menggunakan software Rapid Miner diperoleh nilai akurasi yaitu sebesar 98.73%. Dimana untuk Class Precision pada prediksi label No sebesar 100% dan Class Precision pada prediksi label Yes sebesar 96.43%.

KESIMPULAN

Berdasarkan pengolahan data dan analisis yang telah dilakukan, penulis sampai pada beberapa kesimpulan penting. Pertama, atribut Ekstra Kulikuler memiliki pengaruh signifikan terhadap prediksi prestasi siswa di SMPN 51 Jakarta, karena siswa yang mengikuti kegiatan ekstrakurikuler cenderung berprestasi. Kedua, atribut Alpha juga memiliki pengaruh yang cukup besar, menunjukkan bahwa kehadiran siswa di sekolah memainkan peran penting dalam menentukan absensi siswa yang baik.

Selanjutnya, prediksi prestasi siswa dapat diidentifikasi menggunakan metode Algoritma C4.5 dan Pohon Keputusan yang diolah melalui perangkat lunak Rapid Miner. Hasilnya menunjukkan bahwa atribut "Ekstra Kulikuler" menjadi atribut paling penting dalam proses prediksi, diikuti oleh atribut "Alpha" dengan gain tertinggi. Atribut "Alpha" kemudian digunakan sebagai node akar terakhir dalam pohon keputusan, dengan nilai di atas 0 menunjukkan siswa tidak berprestasi ("NO") dan nilai di bawah atau sama dengan 0 menunjukkan siswa berprestasi.

Akhirnya, metode Algoritma C4.5 yang diterapkan menggunakan Rapid Miner menghasilkan akurasi yang tinggi, mencapai 98,73%, sementara perhitungan manual hanya mencapai 34,18% untuk presentase "YES". Hal ini menunjukkan bahwa metode ini berdampak signifikan dalam memprediksi prestasi siswa di SMPN 51 Jakarta.

REFERENSI

- Handrianto, Y., & Farhan, M. (2019). C.45 Algorithm for Classification of Causes of Landslides. *Sinkron*, 4(1), 120. <https://doi.org/10.33395/sinkron.v4i1.10154>
- Hermawan, I. (2019). *Metodologi Penelitian Pendidikan Kuantitatif, Kualitatif Dan Mixed Method* (C. S. Rahayu (ed.); Cetakan Pe). Hidayatul Quran Kuningan.
- Kusrini, & Luthfi, E. T. (2019). *Algoritma Data Mining* (T. A. Prabawati (ed.)). CV. ANDI OFFSET.
- Liantoni, F. (2022). *Penerapan Data Mining* (E. Setiawan & R. R. Mardiana (eds.)). CV. EUREKA MEDIA AKSARA.
- Muhamad, Windarto, A. P., & Suhada. (2019). Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 1–8. <https://doi.org/10.30865/komik.v3i1.1688>

- Najib, A., Nurcahyono, D., & Setiawan, R. P. P. (2019). Klasifikasi Diagnosa Penyakit Diabetes Mellitus (Dm) Menggunakan Algoritma C4.4. *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, 11(2), 47. <https://doi.org/10.46964/justti.v11i2.153>
- Nurjana, A., Windarto, A. P., & Qurniawan, H. (2022). Implementasi data mining dalam memprediksi prestasi siswa dengan algoritma c4.5. *SmartEDU*, 1(4), 171–180. <https://ejournal.abivasi.id/index.php/SmartEDU/article/view/45>
- Pambudi, R. H., Setiawan, B. D., & Indriati. (2018). Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(7), 2637–2643. <http://j-ptiik.ub.ac.id>
- Ramadhani, R., & Hendriyani, Y. (2021). *Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma Decision Tree (Studi Kasus : SMKN 2 Padang) (Studi Kasus : SMKN 2 Padang) (Font menggunakan Times New Roman berukuran 12 point spasi tunggal) ABSTRAK Pengujian ini bertujuan untuk me.* 9(3). <https://ejournal.unp.ac.id/index.php/voteknika/article/view/112633>
- Rosyid, M. Z., Mustajab, & Abdullah, A. R. (2020). *Prestasi Belajar* (Rofiqi (ed.); 2 Cetakan). Literasi Nusantara.
- Sugiyono. (2015). *Metode Penelitian Kuantitatif Kualitatif Dan R&D.pdf* (8th ed.). Alfabeta.
- Wahono, H., & Riana, D. (2020). Prediksi Calon Pendorong Darah Potensial Dengan Algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 7. <https://doi.org/10.30865/jurikom.v7i1.1953>
- Yendrizal. (2022). *Monograf Algoritma C4.5 Pada Teknik Klasifikasi Penyusutan Volume Pupuk* (E. Suardi (ed.)). CV. AZKA PUSTAKA.